JNeurosci
THE JOURNAL OF NEUROSCIENCE

Research Articles: Systems/Circuits

Inverted encoding models of human population response conflate noise and neural tuning width

Taosheng Liu[1], Dylan Cable[2] and Justin L. Gardner[2]

[1]*Department of Psychology, Michigan State University, East Lansing, MI, USA*
[2]*Department of Psychology, Stanford University, Stanford, CA, USA*

DOI: 10.1523/JNEUROSCI.2453-17.2017

Received: 25 August 2017

Revised: 8 November 2017

Accepted: 10 November 2017

Published: 22 November 2017

Author contributions: T.L. and J.L.G. designed research; T.L. and J.L.G. performed research; T.L., D.C., and J.L.G. contributed unpublished reagents/analytic tools; T.L., D.C., and J.L.G. analyzed data; T.L. and J.L.G. wrote the paper.

Conflict of Interest: The authors declare no competing financial interests.

We owe a debt of gratitude to Kang Cheng (1962-2016) for his spirit of rigorous scientific inquiry and unfailing personal encouragement that has been, and continues to be, a guiding force of our research. We thank Kenji Haruhana and members of the Support Unit for Functional Magnetic Resonance Imaging for assistance in data collection and Toshiko Ikari for administrative assistance. We also thank John Serences for helpful comments on an earlier version of the manuscript. This work was supported by funding from the RIKEN Brain Science Institute, from Stanford's BioX Undergraduate Summer Research Program and VPUE Faculty Grant for Undergraduate Research to D.C., from the NIH to T.L. (R01 EY022727), from Grants-in-Aid for Scientific Research 24300146, Japanese Ministry of Education, Culture, Sports, Science and Technology and the Hellman Faculty Scholar Fund to J.L.G.

Correspondence can be addressed to either Taosheng Liu (tsliu@msu.edu) or Justin Gardner (jlg@stanford.edu).

Cite as: J. Neurosci ; 10.1523/JNEUROSCI.2453-17.2017

Inverted encoding models of human population response conflate noise and neural tuning width

Taosheng Liu[1], Dylan Cable[2] and Justin L. Gardner[2]

[1]Department of Psychology, Michigan State University, East Lansing, MI, USA

[2]Department of Psychology, Stanford University, Stanford, CA, USA

Correspondence can be addressed to either Taosheng Liu (tsliu@msu.edu) or Justin Gardner (jlg@stanford.edu).

Acknowledgements: We owe a debt of gratitude to Kang Cheng (1962-2016) for his spirit of rigorous scientific inquiry and unfailing personal encouragement that has been, and continues to be, a guiding force of our research. We thank Kenji Haruhana and members of the Support Unit for Functional Magnetic Resonance Imaging for assistance in data collection and Toshiko Ikari for administrative assistance. We also thank John Serences for helpful comments on an earlier version of the manuscript. This work was supported by funding from the RIKEN Brain Science Institute, from Stanford's BioX Undergraduate Summer Research Program and VPUE Faculty Grant for Undergraduate Research to D.C., from the NIH to T.L. (R01 EY022727), from Grants-in-Aid for Scientific Research 24300146, Japanese Ministry of Education, Culture, Sports, Science and Technology and the Hellman Faculty Scholar Fund to J.L.G.

Abstract

Channel encoding models offer the ability to bridge different scales of neuronal measurement by interpreting population responses, typically measured with BOLD imaging in humans, as linear sums of groups of neurons (channels) tuned for visual stimulus properties. Inverting these models to form predicted channel responses from population measurements in humans seemingly offers the potential to infer neuronal tuning properties. Here, we test the ability to make inferences about neural tuning width from inverted encoding models. We examined contrast invariance of orientation selectivity in human V1 (both sexes) and found that inverting the encoding model resulted in channel response functions that became broader with lower contrast, thus, apparently, violating contrast invariance. Simulations showed that this broadening could be explained by contrast-invariant single-unit tuning with the measured decrease in response amplitude at lower contrast. The decrease in response lowers the signal-to-noise ratio of population responses that results in poorer population representation of orientation. Simulations further showed that increasing signal-to-noise makes channel response functions less sensitive to underlying neural tuning width, and in the limit of zero noise will reconstruct the channel function assumed by the model regardless of the bandwidth of single-units. We conclude that our data are consistent with contrast invariant orientation tuning in human V1. More generally, our results demonstrate that population selectivity measures obtained by encoding models can deviate substantially from the behavior of single-units because they conflate neural tuning width and noise and are therefore better used to estimate the uncertainty of decoded stimulus properties.

42 **Significance Statement**

43 It is widely recognized that perceptual experience arises from large populations of neurons, rather than
44 a few single-units. Yet, much theory and experiment has examined links between single-units and
45 perception. Encoding models offer a way to bridge this gap by explicitly interpreting population
46 activity as the aggregate response of many single neurons with known tuning properties. Here we use
47 this approach to examine contrast invariant orientation tuning of human V1. We show with experiment
48 and modeling that due to lower signal-to-noise, contrast-invariant orientation tuning of single-units
49 manifests in population response functions that broaden at lower contrast, rather than remain contrast-
50 invariant. These results highlight the need for explicit quantitative modeling when making a reverse-
51 inference from population response profiles to single-unit responses.

52

53

Introduction

Bridging knowledge derived from measurements at different spatial and temporal scales is a significant challenge for understanding the link between neural activity and behavior. While much work has focused on linking single-unit measurements to behavior, there is increasing recognition of the importance of population-scale representations (Benucci et al., 2009; Graf et al., 2011; Churchland et al., 2012; Mante et al., 2013; Fusi et al., 2016). In human neuroscience, these bridging challenges are even more severe as many of the core building blocks of knowledge learned from invasive animal experiments are difficult to verify and replicate in humans. It is therefore often unknown whether basic phenomena from the single-unit literature are applicable to humans, let alone how these phenomena will manifest at the larger scale of population activity that is typically interrogated by non-invasive measurement of the human brain.

Recently, an encoding model approach has proven useful in the analysis of large scale population activity measured by functional imaging (Naselaris et al., 2011; Serences and Saproo, 2012) and offers the promise of bridging knowledge from different species and scales of measurements. Encoding models build off of fundamental results in visual physiology, by encoding complex stimuli in lower dimensional representations such as receptive field or channel models. The assumption is that if these neural representations are operative in human cortex, then large-scale measurements of activity represent the aggregated responses of these basic neural operations. For example, a channel encoding model (Brouwer and Heeger, 2009, 2013) has been used to examine continuous stimulus dimensions such as color or orientation where it is reasonable to expect that there are large groups of neurons, or channels with known selectivity, and that voxel responses can be modeled as linear combinations of such channels. These channel encoding models have been used to examine responses for orientation, color, direction and speed of motion and somatosensory response to better understand apparent motion (Chong et al., 2015), cross-orientation suppression (Brouwer and Heeger, 2011), normalization

(Brouwer et al., 2015), speeded decision making (Ho et al., 2012), attention (Scolari et al., 2012; Garcia et al., 2013; Saproo and Serences, 2014; Ester et al., 2016), working memory (Ester et al., 2013, 2015), perceptual learning (Byers and Serences, 2014; Chen et al., 2015), biases in motion perception (Vintch and Gardner, 2014), and exercise (Bullock et al., 2016) using both functional imaging and EEG (Garcia et al., 2013; Bullock et al., 2016) measurements. Inverting these models to form predictions of channel response from cortical measurements produces tuned response profiles. The interpretations of these tuned response profiles are encouraging for the effort of bridging across measurements as they have shown results in concordance with expectations from electrophysiological measurement of phenomena such as decision-making reliance on off-target populations (Purushothaman and Bradley, 2005; Scolari et al., 2012) and feature-similarity gain (Treue and Maunsell, 1996; Saproo and Serences, 2014) and response gain (McAdams and Maunsell, 1999; Garcia et al., 2013) modulation effects of attention.

Here we test the ability of the channel encoding model approach to bridge single-unit and population scale measurement by asking whether the well-known property of contrast-invariant orientation tuning is manifest in predicted channel responses from human primary visual cortex. We reasoned that examining whether orientation tuning bandwidth of human cortical population responses change with contrast would provide a good test case for the use of encoding models to bridge measurements, because there is a clear prediction of invariance from single-unit measurements (Sclar and Freeman, 1982). However, contrary to the electrophysiology literature, we found that an encoding model produced channel response functions that increased in bandwidth as contrast was lowered. Computational modeling revealed that these effects can be explained by the reduced signal-to-noise ratio of cortical responses at lower contrast. These results emphasize that bridging different levels of measurement through these analyses requires explicit quantitative statements of how properties of single-units are expected to manifest in population activity.

Materials and Methods

Subjects

Six healthy volunteers (ages 33-42, two female) from the RIKEN Brain Science Institute community participated in the experiment; all had normal or corrected-to-normal vision and were experienced subjects in functional imaging experiments. The study protocol was approved by the RIKEN Functional MRI Safety and Ethics Committee and all subjects gave written consent to experimental procedures in advance to participating in the experiment.

Stimuli

Stimuli were generated using MGL, a set of Matlab routines for implementing psychophysical experiments (http://gru.stanford.edu/mgl). Stimuli were back projected onto a screen using a LCD projector (Silent Vision 6011; Avotec) at a resolution of 800×600 and a refresh rate of 60 Hz. Subjects viewed the screen via an angled mirror attached to the head coil. The projector was gamma corrected to achieve a linear luminance output.

Visual stimuli were sinusoidal gratings (spatial frequency: 0.7 cpd) in a circular aperture (10°), located to the left or the right of a central fixation cross (1°) at an eccentricity of 8°. The gratings were either low (20%) or high (80%) contrast, and could be in one of eight evenly spaced orientations from 0° to 180° (see Figure 1).

Task and Procedures

On each trial, two gratings were presented for 5.12 s, followed by a 3.84 s inter-trial interval. During the grating presentation, the phases of both gratings were updated every 0.2 s. The phase of each grating was randomly chosen from one of 16 uniformly distributed phases from 0 to 2π, and the starting time of the phase update of each grating was randomly determined such that the phase updates

of the two gratings were asynchronous. The phase updates were implemented to reduce retinal adaptation and afterimages. The contrast and orientation of each grating was randomly chosen on each trial such that each combination of contrast (two levels) and orientation (eight levels) was presented three times in each run (48 trials in total). While the inter-trial interval was short which could result in non-linear summation of responses from the previous trial (Boynton, et al., 1996), the trial randomization procedure served to minimize previous trial effects as on average they would come from a random trial type. In addition, a fixation period of 5.12 s preceded each run, making each run 435.2 s in the scanner. Subjects completed 9 runs in the scanner (432 trials in total), which yielded 27 trials per orientation/contrast combination.

While the gratings were presented in the periphery, subjects performed a luminance discrimination task at fixation. On each trial in this task, the fixation cross dimmed for 0.4 s twice, separated by a 0.8 s interval, and subjects had to indicate in which interval the cross appeared darker. The magnitude of dimming was held constant for one interval while the magnitude of dimming in the other interval was controlled by a one-up two-down staircase. Subjects pressed one of two keys (1 or 2) to indicate their response. The fixation task was performed continuously throughout a run and was asynchronous with the display of the grating stimuli. This task was used to control subjects' attention and ensure a steady behavioral state and eye fixation. The independently randomized contrast and orientations of the two gratings on either side also served as an internal check of the fixation quality, as any systematic bias of eye position for one stimulus would not be systematic for the other.

MRI methods.

Imaging was performed with a Varian Unity Inova 4T whole-body MRI system (now Agilent Technologies) located at the RIKEN Brain Science Institute, Saitama, Japan. A volume RF coil (transmit) and a four-channel receive array (Nova Medical) were used to acquire both functional and

147 anatomical images.

148 Each subject first participated in a separate scanning session to obtain their retinotopic maps
149 (see below for more details), using standard procedures. During this session, a high-resolution 3D
150 anatomical T1-weighted volume (MPRAGE; TR, 13 ms; TI, 500 ms; TE, 7 ms; flip angle, 11°; voxel
151 size, 1 × 1 × 1 mm; matrix, 256 × 256 × 180) was obtained, which served as the reference volume to
152 align all functional images. The reference volume was segmented to generate cortical surfaces using
153 Freesurfer (Dale et al., 1999). Subsequently, the anatomy volumes taken at the beginning of each
154 session were registered to the reference volume so that the cortical regions in the functional scans were
155 aligned with the retinotopy. All analyses were performed in the original (non-transformed) coordinates
156 before being mapped to the cortical surface and specific visual regions.

157 During the main experiment, functional images were collected using a T2*-weighted echo-
158 planar-imaging sequence (TR, 1.28 s; TE, 25 ms; flip angle, 45°; sensitivity encoding with acceleration
159 factor of 2). We collected 29 slices at an angle approximately perpendicular to the calcarine sulcus,
160 with resolution of 3 × 3 × 3 mm (field of view, 19.2 × 19.2 cm; matrix size, 64 × 64). The first four
161 volumes in each run were discarded to allow T1 magnetization to reach steady state. In addition, a T1-
162 weighted (MPRAGE; TR, 11 ms; TI, 500 ms; TE, 6 ms; flip angle, 11°; voxel size, 3 × 3 × 3 mm;
163 matrix, 64 × 64 × 64) anatomical image was acquired to be used for co-registration with the high-
164 resolution reference volume collected in the retinotopic session.

165 Various measures were taken to reduce artifacts in functional images. During scanning,
166 respiration was recorded with a pressure sensor, and heartbeat was recorded with a pulse oximeter.
167 These signals were used to attenuate physiological signals in the imaging time series using
168 retrospective estimation and correction in *k* space (Hu et al., 1995).

169 *Retinotopic mapping procedure*

In this separate scanning session, we mapped each subject's occipital visual areas using well-established phase-encoding methods (Sereno et al., 1995; DeYoe et al., 1996; Engel et al., 1997), so only a brief description is provided here. We presented rotating wedges and expanding/contracting rings over multiple runs and averaged runs of the same type. Then a Fourier analysis was applied to the averaged time course to derive the polar angle map and eccentricity map from the wedge and ring data, respectively. Borders between visual areas were defined as phase reversals in the polar angle map of the visual field. The map was visualized on computationally flattened representations of the cortical surface generated by FreeSurfer. For each subject, we could readily define many visual areas, including V1, V2, V3, and hV4. However, we will mainly focus on V1 in this study.

BOLD Data Analysis.

Data were processed and analyzed using mrTools (http://gru.stanford.edu/mrTools) and other custom code in MATLAB (MathWorks, Natick, MA). Preprocessing of function data included head movement correction, linear detrend, and temporal high-pass filtering at 0.01 Hz. The functional images were then aligned to high-resolution anatomical images for each participant, using an automated robust image registration algorithm (Nestares and Heeger, 2000). Functional data were converted to percentage signal change by dividing the time course of each voxel by its mean signal over a run, and data from the 9 scanning runs were concatenated for subsequent analysis.

Voxel selection

We used an event-related (finite impulse response or deconvolution) analysis to select voxels in V1 that responded to the stimulus presentation. Each voxel's time series was fitted with a general linear model with regressors for 16 conditions (2 contrasts × 8 orientations) that modeled the BOLD response in a 25 s window after trial onset. The design matrix was pseudo-inversed and multiplied by the time series to obtain an estimate of the hemodynamic response for each stimulus condition. Because the

193 stimulus was independently randomized in the left and right visual field, we fitted two event-related
194 models for each subject, one based on the stimulus in the left visual field and one based on the stimulus
195 in the right visual field.

196 For each voxel, we also computed a goodness-of-fit measure (r^2 value), which is the amount of
197 variance in the BOLD time series explained by the event-related model (Gardner et al., 2005) . In other
198 words, the r^2 value indicates the degree to which a voxel's time course is modulated by the task events,
199 and hence we can use it to select voxels in V1 that were active during the experiment. We selected
200 voxels whose r^2 values were greater than 0.05, which yielded ~100 voxels in each V1 in each
201 hemisphere. Subsequent analysis focused on imaging data in this subset of V1 voxels. Our results did
202 not vary substantially with the voxel selection criterion. For example, when we varied the r^2 cut-off to
203 select a larger number of V1 voxels (~150), the tuning widths of the channel response function were
204 30.6 and 49.9 deg for the high and low contrast stimulus, respectively. For a smaller number of V1
205 voxels (~65), the tuning widths values were 22.7 and 38.7 deg, respectively.

206 *Channel encoding model*

207 We used a channel encoding model (also referred to as an "encoding model" for brevity below),
208 proposed by Brouwer & Heeger (2009), to characterize the orientation tuning of V1 voxels.
209 Conceptually, the model assumes each voxel's response is some linear combination of a set of
210 channels, each channel having the same bandwidth, but with a different preferred stimulus value. We
211 refer to the tuning functions that specify the channels as "model basis functions", which together span
212 the range of all stimulus values. The intuition is that each voxel's response is due to populations of
213 neurons that are tuned to different stimulus values and the analysis proceeds by trying to determine
214 which combination of these neural populations (channels) are most responsible for a voxel's response.
215 For every stimulus presentation, the ideal response of each channel is calculated based on the stimulus

value and model basis functions. The weights of each channel that best fit each voxel's response in the least-squares sense is determined using linear regression from a training data set. Once these weights are fit, the model can be inverted on a left-out test data set to reconstruct channel responses from observed voxel responses. The average channel responses relative to the actual presented stimulus is called a channel response function.

To use as training and test data for the channel encoding model, we obtained single-trial BOLD responses for each V1 voxel with the following procedure. For each V1 hemisphere, we first averaged the event-related BOLD response (see voxel selection above) across all voxels and conditions, which served as an estimate of the hemodynamic impulse response function (HIRF) in each V1 hemisphere. We then constructed a boxcar function for each individual trial (with the boxcar length equaling the length of the stimulus presentation), and convolved it with the estimated HIRF, to produce a design matrix coding for each individual trial in each condition. The design matrix was then pseudo-inversed and multiplied by the time series to obtain an estimate of the response amplitude for each individual trial in each voxel. We call the set of response amplitudes across all voxels in a V1 hemisphere a response "instance".

We fit the encoding model to the instances with a 5-fold cross-validation scheme, in which 4/5 of the trials were randomly selected to be the training data and the remaining 1/5 of trials constituted the test data. This analysis was performed on instances for the high contrast and low contrast trials separately (216 trials per contrast). Our encoding model consisted of 8 evenly-spaced channels (i.e., model basis functions) from 0 to 180°, with each channel a half-wave rectified sinusoid raised to the power of 7. These basis functions were chosen to approximate single-neuron's orientation tuning function in V1. In the following exposition, we adopted the notation from Brouwer & Heeger (2009). The training instances can be expressed as a $m \times n$ matrix $\mathbf{B_1}$, where m is the number of voxels, and n is number of trials in the training data. We then constructed hypothetical channel outputs given the

stimulus orientation on each trial of the training dataset, which yielded a $k \times n$ matrix $\mathbf{C_1}$, where k is the number of channels (i.e., k=8). Each column in the $\mathbf{C_1}$ matrix represented a set of ideal response to the stimulus orientation on that trial from the eight channels. A weight matrix $\mathbf{W}$ ($m \times k$) relates the observed data $\mathbf{B}$ and the hypothetical responses

$$\mathbf{B_1} = \mathbf{W}\,\mathbf{C_1} \qquad \text{(Equation 1)}$$

Each row of $\mathbf{W}$ represents the relative contribution of the eight channels to that voxel's response. The least-square estimate of $\mathbf{W}$ was obtained with the following equation (T indicates the transpose of the matrix):

$$\hat{\mathbf{W}} = \mathbf{B_1}\,\mathbf{C_1}^{T}\,(\mathbf{C_1}\,\mathbf{C_1}^{T})^{-1} \qquad \text{(Equation 2)}$$

The test instances can be expressed as a $m \times p$ matrix $\mathbf{B_2}$, where p is the number of trials in the test data. The estimated channel response to each test stimulus ($\hat{\mathbf{C}}_2$) can then be estimated using the weights $\mathbf{W}$:

$$\hat{\mathbf{C}}_2 = (\hat{\mathbf{W}}^{T}\,\hat{\mathbf{W}})^{-1}\,\hat{\mathbf{W}}^{T}\,\mathbf{B_2} \qquad \text{(Equation 3)}$$

$\hat{\mathbf{C}}_2$ is a $k \times p$ matrix, with each column representing each channel's response to the stimulus on that test trial. The columns of $\hat{\mathbf{C}}_2$ were circularly shifted such that the channel aligned to the test stimulus on that trial was centered in the orientation space. The shifted columns were then averaged to obtain a mean channel response. This procedure was repeated in each fold of the cross-validation, and the mean channel responses from each fold were further averaged to obtain what we will refer to as a "channel response function". For each V1 hemisphere in each participant, we obtained channel response functions for both the contralateral and ipsilateral stimulus, separately for the high contrast and low contrast conditions.

In addition to the channel response function, we also calculated a goodness-of-fit measure of

the encoding model. In each cross-validation fold, after we obtained estimates of the channel weights $\hat{\mathbf{W}}$, we also constructed another set of hypothetical channel outputs given the actual stimulus orientation in the test data, $\mathbf{C_2}$ (note that this is a matrix similar to $\mathbf{C_1}$ but for the test trials, and differs from $\hat{\mathbf{C}}_2$ which is the *estimated* channel responses based on the voxel responses). The predicted voxel response can be obtained via:

$$\mathbf{B_{2pred}} = \hat{\mathbf{W}}\,\mathbf{C_2} \qquad \text{(Equation 4)}$$

To the extent that the encoding model provides a good fit to the data, the predicted response $\mathbf{B_{2pred}}$ ($m \times p$) should be similar to the observed response, $\mathbf{B_2}$. We can thus calculate the amount of variance explained by the model as

$$r^2 = 1 - \frac{\sum(B_{2pred} - B_2)^2}{\sum(B_2 - \bar{B}_2)^2} \qquad \text{(Equation 5)}$$

$\overline{\boldsymbol{B_2}}$ is the mean voxel responses across voxels and trials, and the summation was performed across voxels and trials (i.e., all values in the $m \times p$ matrix). We calculated r^2 from each fold of the cross validation and averaged them across the folds to obtain a single measure of the goodness-of-fit, for each contrast. Note this r^2 is different from the r^2 that represents the goodness-of-fit of the event-related model (see above section *Voxel selection*). In the remainder of this report, we will focus on this r^2 value that indexes the goodness-of-fit of the channel encoding model.

Quantifying the channel response function

We fitted a circular bell shaped function (von Mises) to the channel response function

$$y = y_0 + Ae^{\kappa cos(x-\mu)} \qquad \text{(Equation 6)}$$

Where κ is the concentration parameter which controls the width of the function, μ is the mean, x is the orientation, A is the amplitude parameter, and y_0 is the baseline. Thus there were four free parameters of the fit: baseline (y_0), amplitude (A), mean (μ), and concentration (κ). Because orientation is on [0 π],

284 whereas the von Mises function spans the interval [0 2π], orientation values were multiplied by 2
285 during the fit, after which the fitting results were scaled back to [0 π]. Fitting was performed using a
286 non-linear least square method, as implemented in MATLAB. To ease the interpretation of the results,
287 we report the half-width-at-half-maximum instead of the concentration parameter, κ, because the latter
288 is inversely related to the variance.

289 *Linking neuronal tuning to channel response: a computational model*

290 We implemented a computational model that links neuronal tuning to channel response
291 functions, using the assumptions underlying the channel encoding model. Given the channel response
292 function is a highly derived statistic, this modeling effort was used to clarify how various assumptions
293 of neural tuning and signal-to-noise would manifest in channel response functions. Specifically, we
294 used this model to fit our observed data and simulate other scenarios to test the validity of the encoding
295 model. The schematic of the model is outlined in Figure 3. The model contains 100 V1 voxels,
296 comparable to the actual number of voxels used in our data analysis (see above). Each voxel is assumed
297 to contain neurons tuned to all orientations, whose tuning functions are described by von Mises
298 functions (see above). The preferred orientation (μ) are evenly distributed across all possible
299 orientations in 1° increment (left column, "Neural tuning functions"), forming 180 classes of neurons.
300 The width of the neural tuning function is specified by the concentration parameter, κ, which is the
301 same for all neurons but can be manipulated across different simulations. The area under the orientation
302 tuning function for neurons was normalized to one so that average firing rate for each neuron would not
303 vary with tuning width. For each voxel, a weight vector, $\boldsymbol{w}$, was generated by randomly sampling 180
304 numbers from [0 1]. This weight vector specifies how much each class of neurons contributes to the
305 voxel's response. For an input stimulus with an orientation, θ, the response of each neuron is calculated
306 according to its tuning function (see Eq. 6). For each voxel, the responses of individual neurons are
307 multiplied by the weight vector, $\boldsymbol{w}$, and then summed to arrive at a predicted response (middle column,

"Random weights"). To calculate the final voxel response, Gaussian noise with *N(0, σ)*, is then added to this response to simulate physiological and thermal noise in BOLD measurements. Thus, each voxel's response is determined by neuronal tuning width (κ), weight vector (***w***), and noise (σ). Note ***w*** is randomly generated for each voxel in each simulation, whereas κ and σ are parameters that we examined systematically in several simulations (see Results).

We simulated experiments with the same basic set-up as our empirical study: 8 possible orientations, with each orientation shown 27 times (trials). For each trial, we obtained a vector of voxel responses (instances), calculated as above. Then all the trial instances were subjected to the same analysis as the real data described above, i.e., cross-validation in which 4/5 of data were used to obtain the channel weights and 1/5 of the data were used to obtain channel response functions. We used the same exact code to analyze the synthetic and real data.

Computation of the posterior distributions

We computed the probability of different stimulus values given the test data, i.e., the posterior distribution, using a technique from van Bergen et al. (2015). The method begins with finding the weights of the channel encoding model as above. After removing the signal due to the encoding model, a noise model is fit to the residual response. The noise model assumes that each individual voxel's variability is Gaussian with one component which is independent among voxels and another that is shared across all voxels. Each of the channels is modeled to have independent, identically distributed Gaussian noise. This leads to a covariance matrix for the noise as follows:

$$\mathbf{\Omega} = \rho\boldsymbol{\tau}\boldsymbol{\tau}^T + (1-\rho)\,\mathbf{I} \blacklozenge \boldsymbol{\tau}\boldsymbol{\tau}^T + \sigma^2\hat{\mathbf{W}}\hat{\mathbf{W}}^T \qquad \text{(Equation 7)}$$

Where $\mathbf{\Omega}$ is the noise model's covariance matrix, $\mathbf{I}$ is the identity matrix, ◆ denotes element-wise multiplication, $\boldsymbol{\tau}$ is a vector containing each voxel's independent standard deviation, ρ is a scalar between 0 and 1 which controls the amount of shared variability among voxels, σ is the standard

deviation of each channel and $\hat{\mathbf{W}}$ is the estimated weight matrix from Equation 2. This noise model with parameters, τ, ρ and σ is fit via maximum likelihood estimation to the residual and can be used to compute the probability of generating any particular response given a stimulus value. Inversion of this equation using Bayes' rule and a flat prior allows one to compute the probability of any stimulus value given a response—the posterior distribution. For a derivation and detailed explanation see (van Bergen et al., 2015). All analysis followed the same 5-fold cross-validation scheme used for the encoding model by which 4/5 of the data were used to fit the model weights and noise model parameters and the 1/5 left out data were used to compute the posterior distribution using Bayes' rule. Results are shown averaged across all five left-out folds.

341 **Results**

342 We measured BOLD responses from retinotopically defined V1 to oriented sinusoidal gratings
343 (Fig 1) and used the resulting data to train and test a channel encoding model. We first report average
344 (univariate) activity across voxels. While responses averaged across subjects and voxels in V1 did not
345 show any apparent selectivity for orientation (Fig 2A), they were, as expected (Boynton et al., 2012),
346 greater for higher contrast compared to low contrast contralateral stimuli (compare red vs. yellow).
347 Consistent with the laterality of V1, responses did not vary with contrast of the ipsilateral stimulus
348 (black). The presence of a strong contrast response for the contralateral stimuli but a complete absence
349 of such a response for ipsilateral stimuli also suggest that subjects maintained stable central fixation.

350 Despite this lack of orientation selectivity, channel response functions obtained from the
351 encoding model and averaged across subjects displayed peaked responses at the true orientation of the
352 contralateral stimulus for both high (Fig 2B, red) and low (Fig 2C, yellow) contrast stimuli. We
353 computed r^2, a measure of goodness-of-fit, which showed that the channel encoding model accounted
354 for 31% and 13% (high and low contrast contralateral stimuli, respectively) amount of the variability of
355 the data. The ability to recover these peaked function of orientation is consistent with previous studies
356 using classification approaches (Kamitani and Tong, 2005) and is presumably due to biases in response
357 to orientation which differ for individual voxels but are eliminated when responses are averaged across
358 voxels. As expected from the lateralization of V1, these channel response functions are flat when
359 constructed for the ipsilateral stimulus (black), thus serving as an internal control on the validity of the
360 encoding model approach.

361 However, contrary to the expectations of contrast-invariance, channel response functions were
362 broader for low contrast compared to high contrast contralateral stimuli (Fig 2D, compare yellow to
363 red). Fitting a bell shaped function (von Mises) to the subject-averaged channel response functions

364 revealed that the tuning width went from 25.6 to 42.0 deg (half-width at half-height) as contrast was
365 lowered. Amplitude was also decreased for the low contrast stimuli from 0.27 to 0.23, where 1 would
366 be the ideal height of the channel response function if voxel responses contained noise-free information
367 about stimulus orientation. This pattern of results was also evident in individual subject's channel
368 response functions. 9 out of 12 hemispheres showed a decrease in tuning width as contrast was
369 decreased ($p = 0.013$, $t(11) = -2.60$, one-tailed paired t-test). 10 out of 12 hemispheres had lower
370 amplitude for the low contrast condition ($p = 0.042$, $t(11) = 1.90$, one-tailed paired t-test). We also
371 examined extrastriate areas V2, V3, and hV4 and found similar results. In V2, tuning width went from
372 21.3 to 75.5 deg and amplitude went from 0.25 to 0.19 as contrast was lowered. In V3, tuning width
373 went from 17.8 to 88.6 deg and amplitude went from 0.23 to 0.18 as contrast was lowered. Finally, in
374 hV4, tuning width for high contrast stimuli was 14.2 deg and channel response function was essentially
375 flat for low contrast stimuli. Below we will focus on results from V1, which is best informed by
376 neurophysiological results (Sclar and Freeman, 1982; Skottun et al., 1987; Carandini et al., 1997).

377 While the lack of contrast-invariant channel response functions might imply broader neuronal
378 tuning at low contrast in human visual cortex, we instead considered whether it might be due to the
379 weaker stimulus-driven signal at lower contrast. As noted above, BOLD responses had lower amplitude
380 with lower contrast (Fig 2A). Given that many sources of noise in BOLD are non-neural (e.g.
381 hemodynamic variability and head motion) and thus not expected to vary with signal strength, these
382 lower amplitude responses result in lower signal-to-noise ratio (SNR) of the measurements made with
383 lower contrast. Indeed, the amount of variance accounted for by the encoding model, r^2, was
384 significantly lower for low contrast compared to high contrast stimuli for 12 out 12 hemispheres ($p <$
385 0.001, $t(11) = 5.58$, one-tailed paired t-test). In the extreme case, channel response functions built on
386 responses without any signal, as for the ipsilateral stimulus, are flat. Thus, we reasoned that the lower
387 SNR measurements at low contrast could also result in flatter, i.e. broader, channel response functions.

To test whether reduced SNR at low contrast, rather than changes in neural tuning, could account for the increased channel response bandwidth at low contrast, we built simulations (Fig 3, see Materials and Methods). Briefly, each simulated voxel received randomly weighted responses from simulated orientation-tuned neurons. Different voxels had different weightings of the neuronal responses, thus resulting in weak, but different orientation selectivity across voxels. We added random Gaussian noise to these voxel responses and trained and tested the encoding model using the same procedures as we did for the actual BOLD data. We varied the standard deviation of the added noise (σ) to produce channel response functions that best fit the empirical data (in the least-squares sense) from the high contrast trials (Fig 4A). We then noted that in the empirical data there was a 42.2% decrease in neural response from high to low contrast across voxels (Fig 2A). We therefore decreased neural response by this amount for all the simulated neurons and found that the resulting channel response function to be a reasonable fit for the low contrast data (Fig 4B). Importantly, this good correspondence between model predictions and data was achieved without fitting any parameter, because the only thing we changed in the simulation was to decrease the magnitude of response across all neurons according to the value found from the empirical data. This suggests that reductions in response magnitude, and therefore SNR, are sufficient to produce changes in channel response width commensurate with what we observed.

While tuning width is not expected to change with contrast, what if we had tested a property for which we expect a neural tuning width change? Do channel response functions track changes in neural tuning width? We simulated neural tuning functions from 5 to 40 degrees half-width at half-height as well as a "stick function" which responds only to a single orientation maximally and does not respond to any other orientation and computed channel response functions under different amounts of noise (Fig 5A, cyan to magenta curves represent different neural tuning widths). We found that the resulting channel response functions did indeed track the neural tuning widths, but as the goodness-of-fit r^2,

412 increased (achieved by varying the standard deviation of the added noise, abscissa in Fig 5B), the
413 difference in channel response widths was diminished (note the larger splay of curves on the left vs
414 right side of Fig 5B). Thus, channel response functions can reflect underlying neural tuning widths, but,
415 perhaps counterintuitively, the better the goodness-of-fit of the encoding model, the less difference the
416 neural tuning width makes.

417 To understand why better goodness-of-fit implies worse model discriminability of underlying
418 neural tuning functions, it is important to note that the absolute width of channel response functions is,
419 in the limit of no noise, determined by the basis functions used in the encoding model and not by the
420 neural tuning widths themselves. We simulated two variations of the encoding model in which we
421 varied the model basis function widths. We increased the basis function width by decreasing the
422 exponent on the sinusoidal basis to 3 (Fig 5A) and decreased the basis function width by using stick
423 filters (Fig 5C). For all the simulations, as the goodness-of-fit increases, the recovered channel
424 response functions approach the width of the model basis function (dashed line) rather than the
425 neuronal response width (note that for both the channel response functions and the model basis
426 functions, we use the fitted half-width at half-height as our measure of tuning width and therefore the
427 stick functions do not have infinitely narrow tuning). Given that the encoding model is essentially a
428 linear regression model, this is not an unexpected outcome. Linear weights are being determined to best
429 map the voxel responses into ideal channel responses. As long as the voxel responses are determined
430 by stimulus orientation, then the regression model will be able to recreate exactly any model basis
431 functions that can be formed as linear combinations of the represented orientations

432 The above analysis suggests that while absolute neural tuning width may not be readily
433 determined from the channel response function width, *changes* in tuning width might be meaningful if
434 SNR does not change between conditions. That is, reading vertically for one level of goodness-of-fit in
435 Fig 5A-C, the channel response function width changes systematically as a function of neural tuning

436 width. Might we be able to determine that neural tuning width has changed if we observe data in which
437 we have matched goodness-of-fit of the encoding model? While this does not occur for changes in
438 contrast, this could be the case for potential modulation of tuning width by cognitive factors like
439 attention and learning.

440 However, changes in neural tuning width also result in changes in the model's goodness-of-fit
441 thus complicating the possibility of interpreting changes in channel response width. We simulated
442 different neural tuning widths from 5-45 degrees half-width at half-height (Fig 6, abscissa) for different
443 amounts of additive noise (standard deviation ranging on a logarithmic scale from 0.001 to 1, blue to
444 yellow curves). As can be appreciated by the downward slopes of the curves in Fig 6, as the neural
445 tuning widths get larger, the fit of the encoding model gets worse (lower r^2, ordinate). The reason for
446 this worsened fit is because as neural tuning width gets wider, there is less information about
447 orientation available to fit the model. In the extreme, a flat neural tuning function would result in no
448 orientation specific response and the encoding model would fail to fit the data completely. Each curve
449 in the simulation is what one might expect to measure if neural tuning width is the only variable that
450 changes in the experiment and noise is due mostly to external factors that do not change with
451 conditions. That is, if one expects only a change in neural tuning width, the resulting channel response
452 functions would be expected to have both wider tuning and lower r^2. This pattern of results would make
453 it difficult, if not impossible, to determine whether the changes in tuning were due to decreased SNR,
454 decreased neural tuning width, or some combination of both.

455 These empirical and simulation results suggest caution in interpreting changes in channel
456 response functions because they could be due to either or both changes in neural tuning width and
457 changes in signal strength. What better ways are there for interpreting these functions? A recent report
458 (van Bergen et al., 2015) proposes to transform these channel response functions into posterior
459 distributions that show the probability of different stimulus values given the measured response. We

460 applied the same analysis to our channel response functions, by estimating the distribution of noise in
461 the voxels and model to determine the probability of measuring responses given any oriented stimulus
462 and then applying Bayes' rule with a flat prior to obtain the probability of various stimuli given the
463 responses we measured in a left-out validation set of data (see Materials and Methods for details, Fig
464 7). This resulted in posterior distributions that were peaked around the actual orientation for the
465 contralateral stimulus (red and orange curves), but flat for the ipsilateral stimulus (black curves), thus
466 replicating the results on channel responses. This transformation of the results into posterior
467 distributions allows for a more straight-forward interpretation of the encoding model approach —
468 encouraging interpretation in terms of the certainty by which a given neural response tells us what the
469 stimulus was, rather than what it implies about underlying neural tuning functions.

470

471

Discussion

Using an encoding model approach we built channel response functions for orientation and found that, unlike contrast-invariant single-units, they were broader at lower contrast. Simulations showed that this effect could be fully explained by the measured decrease in overall neural response between high and low contrast, which results in lower signal-to-noise ratio. As signal-to-noise is increased, channel response functions become narrower, until, in the limit of no noise, they approximate the shape of the model basis function (and not the underlying neural tuning function). While changes in underlying neural selectivity in our model could be reflected in channel response functions, our results demonstrate that changes in channel response functions do not necessarily reflect changes in underlying neural selectivity.

Orientation selectivity of single-unit responses have been shown to be invariant to image contrast (Sclar and Freeman, 1982; Skottun et al., 1987; Carandini et al., 1997), suggestive of a general neural computational mechanism (Carandini and Heeger, 2012) by which visual perception can remain relatively unaffected by differences in visibility of stimuli. Despite this central theoretic importance, obtaining non-invasive measurement of selectivity bandwidth from human cortex has been technically difficult because orientation selectivity is organized into cortical columns (Hubel and Wiesel, 1962, 1968; Blasdel and Salama, 1986; Bonhoeffer and Grinvald, 1991), much smaller than the typical spatial resolution of blood-oxygen level dependent (Ogawa et al., 1990, 1992) measurements. While direct measurements of such columnar structures in humans has been achieved (Cheng et al., 2001; Sun et al., 2007; Yacoub et al., 2007, 2008), multivariate analysis using pattern classification approach to decode orientation and motion direction (Haynes and Rees, 2005; Kamitani and Tong, 2005, 2006) from distributed activity patterns has become a more common approach (Norman et al., 2006). However, this classification approach generally produces a categorical outcome, for example, which of two orientations was more likely to have resulted in the measured response pattern and thus is not typically

used for probing selectivity bandwidth of neural representations. The encoding model approach allows one to reconstruct a response profile for a stimulus that has a tuning bandwidth that can be inspected across different contrasts.

While we found an increase in the bandwidth of channel response functions for lower contrast stimuli from human V1, this increase could be fully accounted for by the measured reduction of response amplitude due to contrast, thus reconciling our data with contrast-invariant orientation tuning. We recognize that contrast invariance at the population level as measured with BOLD is not guaranteed even if single-unit spiking responses display contrast-invariance. Systematic relationships between contrast-sensitivity and selectivity for orientation could result in population responses changing selectivity with contrast. For example, if the least orientation-selective neurons saturate their responses at lower contrast than the most selective neurons, then population response would become more selective as contrast increases because population response would be dominated by the most selective neurons. However, no such systematic relationship has been observed and population spiking responses appear contrast-invariant in cats (Busse et al., 2009), consistent with our results. Furthermore, BOLD measurements may be better correlated with local field potentials than spiking activity (Logothetis et al., 2001), which could also result in deviations of BOLD population measures of contrast invariance and spiking activity of neurons. If BOLD measures are sensitive to sub-threshold, synaptic activity that can contribute to local field potentials, broadening of channel response functions that we observed could be reflective of sub-threshold activity, if such activity is not contrast-invariant. However, intracellular measurements of membrane potentials show that selectivity does not broaden at lower contrast, in fact, selectivity is slightly increased at low contrast (Finn et al., 2007), consistent with our interpretation that channel response function broadening at low contrast is due to reduction in signal-to-noise.

Given our results, bridging effects of attention on single units with effects uncovered using

encoding models of functional imaging measurements (Sprague et al., 2015) may be similarly complicated as bridging contrast invariance effects. Single unit studies have suggested that neurons change gain, not selectivity bandwidth (McAdams and Maunsell, 1999; David et al., 2008) with spatial attention, a key finding that has shaped our understanding of neural mechanisms of attention (Carrasco, 2011; Ling et al., 2015). In human population measurements, improved orientation encoding has been found when orientation (but not contrast) is task relevant (Jehee et al., 2011; Ling et al., 2015). While it would be of interest to know whether these population effects of attention reflect differences in neural tuning bandwidth, selective attention, like image contrast, also modulates response amplitudes (Brefczynski and DeYoe, 1999; Gandhi et al., 1999; Kastner et al., 1999; Kastner and Ungerleider, 2000; Reynolds and Chelazzi, 2004) and thus is expected to improve signal-to-noise ratio for population measures. Similarly to contrast effects, attention should be expected to bias channel response functions toward a narrower tuning even if neural tuning bandwidth does not change.

A similar disconnect between single-unit and population measures impacts even simpler measures of cortical response that do not require multivariate approaches. Contrast sensitivity can be directly imaged because single-units monotonically increase response with contrast (Albrecht and Hamilton, 1982; Sclar et al., 1990; Busse et al., 2009) resulting in a population response that also monotonically increases (Tootell et al., 1998; Boynton et al., 1999; Logothetis et al., 2001; Avidan and Behrmann, 2002; Olman et al., 2004; Gardner et al., 2005). Spatial attention has generally been shown to shift contrast response vertically upward when measured with functional imaging (Buracas and Boynton, 2007; Li et al., 2008; Murray, 2008; Pestilli et al., 2011; Hara and Gardner, 2014), which appears to be different from the variety of effects from contrast-gain to response-gain reported for single-units (Reynolds et al., 2000; Martínez-Trujillo and Treue, 2002; Williford and Maunsell, 2006; Lee and Maunsell, 2010; Pooresmaeili et al., 2010; Sani et al., 2017). Consideration of normalization and the size of the attention field relative to stimulus-driven responses can give rise to effects that can

account for single-unit responses and EEG measures (Reynolds and Heeger, 2009; Itthipuripat et al., 2014). But, predictions of this normalization model of attention may differ for single-units and population measures as different neurons in a population can be exposed to different balance of attention field and stimulus-drive, giving rise to additive shifts when considered as a population (Hara et al., 2014). Relatedly, response gain changes may also manifest as additive shifts when directly examining voxel feature selectivity (Saproo and Serences, 2010).

While neural tuning width can be reflected in channel response functions, neural tuning width and signal-to-noise changes are intertwined, making it hard to disentangle their effects. For example, one might examine conditions in which signal-to-noise is matched and then hope to attribute changes in channel response function bandwidth solely to changes in neural tuning bandwidth. However, our simulations show signal-to-noise measures such as the variance accounted for by the encoding model (r^2), covaries with neural tuning width. As neural tuning width broadens there is less modulation of voxel response with orientation and thus the encoding model shows a decrease in r^2. Therefore, even pure changes in neural tuning width would result in conditions with lower r^2, making it hard to attribute changes in channel response functions solely to changes in neural tuning width.

The results of our simulation are agnostic to the source of selectivity for orientation in voxels. One possible source of orientation information are the irregularities of columnar organization which could give rise to small, idiosyncratic biases in voxels (Boynton, 2005; Swisher et al., 2010). However, large scale biases for cardinal (Furmanski and Engel, 2000; Sun et al., 2013) and radial (Sasaki et al., 2006) orientations have been reported, and these biases have been shown to be an important source of information to drive classification (Freeman et al., 2011, 2013; Beckett et al., 2012; Wang et al., 2014; Larsson et al., 2016), but see (Alink et al., 2013; Pratte et al., 2016). Large-scale biases may result from vascular (Gardner, 2010; Kriegeskorte et al., 2010; Shmuel et al., 2010) or stimulus aperture (Carlson, 2014) related effects. Our simulations do not require, or exclude, any topographic arrangement of

biases. Regardless of the source of orientation bias, channel response function widths would be expected to broaden as signal-to-noise decreases.

More generally, our results suggest a "reverse-inference" problem (Aguirre, 2003; Poldrack, 2006) when interpreting outputs from inverted encoding models. Forward encoding from hypothetical neural responses to population activity is a powerful tool, but reversing this process to infer about neural responses is problematic when there is not a one-to-one mapping between single-unit and population measures. Consequently, this reverse-inference problem is not restricted to channel encoding models, but will occur for other encoding model approaches such as population receptive fields (Dumoulin and Wandell, 2008) or Gabor wavelet pyramids (Kay et al., 2008), if one were to invert these models to infer properties of the underlying neural responses. For contrast and orientation, both increases in response amplitude and neural selectivity can result in narrower bandwidth of the channel response functions so reverse inference requires taking both into account. Regardless of which neural change has occurred, read-out of these responses, be they in the brain or from external measurement, will have less certainty about what stimulus has caused those responses. Techniques that represent the output of encoding models as posterior distributions (van Bergen et al., 2015) offer a straightforward interpretation of the uncertainty in determining stimulus properties from cortical responses.

Figure legends

Figure 1. Schematics of the experiment. Low and high contrast gratings of eight possible orientations were presented, with contrast and orientation independently randomized for each visual field. A luminance change-detection task was performed at the fixation to control subjects' fixation and state of attention.

Figure 2. **(A)** Group-averaged mean BOLD response across V1 for each orientation, separately for the contralateral and ipsilateral high- and low- contrast stimuli. **(B)** Group-averaged channel response functions from V1 to a high contrast grating. contra: response calculated for the contralateral stimulus; ipsi: response calculated for the ipsilateral stimulus. **(C)** Same as B except for low contrast grating. **(D)** Group-averaged channel response functions to the low and high contrast grating in the contralateral visual field (symbols, same as the contralateral response in B and C). Solid lines are best fitting von Mises functions to each contrast level. Error bars in these graphs are standard error of the mean (s.e.m.) across participants.

Figure 3. Schematic of the model linking neuronal response to channel response. Each voxel (right column) received randomly weighted responses from orientation-tuned neurons (left column). After weighting and summing, random Gaussian noise was added to obtain simulated voxel responses (see text for more details).

Figure 4. Model predictions of empirical channel response functions. **(A)** Empirical channel response function for contralateral high contrast stimuli (red symbols, same data as in Fig 2B) were fit by the computational model, with the best-fitting channel response shown in black symbols and lines. **(B)** Empirical channel response function for contralateral low contrast stimuli (red symbols, same data as in Fig 2C) and the channel response from the same model used in A (black symbols and lines), except that the neuronal response amplitude was reduced. Error bars are standard error across subjects and

609 hemispheres.

610 ***Figure 5***. Model simulations of how channel response function varies with neural tuning, signal-to-
611 noise ratio, and model basis function. Each panel uses a different model basis function (shown in the
612 top graph) to derive channel responses from synthetic data generated with different combinations of
613 signal-to-noise (r^2, x-axis) and neural tuning width (colored lines). The width of the channel response
614 function is plotted on the y-axis. Horizontal dashed lines indicate the width of the model basis function.

615 ***Figure 6***. Model simulations of how goodness-of-fit of the encoding model (r^2) varies with neural
616 tuning width and noise level in the synthetic data. Different colors represent different amount of
617 Gaussian noise added to the simulated neural response.

618 ***Figure 7***. Posterior distributions from the Bayesian analysis. These functions represent the probability
619 that a given stimulus (measured by the offset from the true orientation, x-axis) caused the observed
620 BOLD response. Posterior distributions for high and low contrast contralateral stimuli are shown in red
621 and yellow, respectively, whereas posterior distributions for ipsilateral stimuli are shown in black.
622 Shaded region represents the standard error over subjects and hemispheres.

623

624

625 **References**

626 Aguirre GK (2003) Functional Imaging in Behavioral Neurology and Cognitive Neuropsychology.
627 Behav Neurol Cogn Neuropsychol:35–46.

628 Albrecht DG, Hamilton DB (1982) Striate cortex of monkey and cat: contrast response function. J
629 Neurophysiol 48:217–237.

630 Alink A, Krugliak A, Walther A, Kriegeskorte N (2013) fMRI orientation decoding in V1 does not
631 require global maps or globally coherent orientation stimuli. Front Psychol 4:493.

632 Avidan G, Behrmann M (2002) Correlations between the fMRI BOLD signal and visual perception.
633 Neuron 34:495–497.

634 Beckett A, Peirce JW, Sanchez-Panchuelo R-M, Francis S, Schluppeck D (2012) Contribution of large
635 scale biases in decoding of direction-of-motion from high-resolution fMRI data in human early
636 visual cortex. Neuroimage 63:1623–1632.

637 Benucci A, Ringach DL, Carandini M (2009) Coding of stimulus sequences by population responses in
638 visual cortex. Nat Neurosci 12:1317–1324.

639 Blasdel GG, Salama G (1986) Voltage-sensitive dyes reveal a modular organization in monkey striate
640 cortex. Nature 321:579–585.

641 Bonhoeffer T, Grinvald A (1991) Iso-orientation domains in cat visual cortex are arranged in pinwheel-
642 like patterns. Nature 353:429–431.

643 Boynton GM (2005) Imaging orientation selectivity: decoding conscious perception in V1. Nat
644 Neurosci 8:541–542.

645 Boynton GM, Demb JB, Glover GH, Heeger DJ (1999) Neuronal basis of contrast discrimination.

Vision Res 39:257–269.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 16:4207–4221.

Boynton GM, Engel SA, Heeger DJ (2012) Linear systems analysis of the fMRI signal. Neuroimage 62:975–984.

Brefczynski JA, DeYoe EA (1999) A physiological correlate of the “spotlight” of visual attention. Nat Neurosci 2:370–374.

Brouwer GJ, Arnedo V, Offen S, Heeger DJ, Grant AC (2015) Normalization in human somatosensory cortex. J Neurophysiol 114:2588–2599.

Brouwer GJ, Heeger DJ (2009) Decoding and reconstructing color from responses in human visual cortex. J Neurosci 29:13992–14003.

Brouwer GJ, Heeger DJ (2011) Cross-orientation suppression in human visual cortex. J Neurophysiol 106.

Brouwer GJ, Heeger DJ (2013) Categorical clustering of the neural representation of color. J Neurosci 33:15454–15465.

Bullock T, Elliott JC, Serences JT, Giesbrecht B (2016) Acute Exercise Modulates Feature-selective Responses in Human Cortex. J Cogn Neurosci:1–14.

Buracas GT, Boynton GM (2007) The effect of spatial attention on contrast response functions in human visual cortex. J Neurosci 27:93–97.

Busse L, Wade AR, Carandini M (2009) Representation of Concurrent Stimuli by Population Activity in Visual Cortex. Neuron 64:931–942.

667 Byers A, Serences JT (2014) Enhanced attentional gain as a mechanism for generalized perceptual
668 learning in human visual cortex. J Neurophysiol 112:1217–1227.

669 Carandini M, Heeger D (2012) Normalization as a canonical neural computation. Nat Rev Neurosci:1–
670 12.

671 Carandini M, Heeger DJ, Movshon JA (1997) Linearity and normalization in simple cells of the
672 macaque primary visual cortex. J Neurosci 17:8621–8644.

673 Carlson T (2014) Orientation Decoding in Human Visual Cortex: New Insights from an Unbiased
674 Perspective. J Neurosci 34:8373–8383.

675 Carrasco M (2011) Visual attention: The past 25 years. Vision Res 51:1484–1525.

676 Chen N, Bi T, Zhou T, Li S, Liu Z, Fang F (2015) Sharpened cortical tuning and enhanced cortico-
677 cortical communication contribute to the long-term neural mechanisms of visual motion
678 perceptual learning. Neuroimage 115:17–29.

679 Cheng K, Waggoner R a, Tanaka K (2001) Human ocular dominance columns as revealed by high-
680 field functional magnetic resonance imaging. Neuron 32:359–374.

681 Chong E, Familiar AM, Shim WM (2015) Reconstructing representations of dynamic visual objects in
682 early visual cortex. Proc Natl Acad Sci U S A 113:1453–1458.

683 Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, Shenoy K V.
684 (2012) Neural Population Dynamics During Reaching. Nature 487:1–20.

685 Dale AM, Fischl B, Sereno MI (1999) Cortical Surface-Based Analysis: I. Segmentation and Surface
686 Reconstruction. Neuroimage 9:179–194.

687 David S V, Hayden BY, Mazer JA, Gallant JL (2008) Attention to stimulus features shifts spectral

688 tuning of V4 neurons during natural vision. Neuron 59:509–521.

689 DeYoe EA, Carman GJ, Bandettini P, Glickman S, Wieser J, Cox R, Miller D, Neitz J (1996) Mapping
690 striate and extrastriate visual areas in human cerebral cortex. Proc Natl Acad Sci U S A 93:2382–
691 2386.

692 Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex.
693 Neuroimage 39:647–660.

694 Engel SA, Glover GH, Wandell BA (1997) Retinotopic organization in human visual cortex and the
695 spatial precision of functional MRI. Cereb Cortex 7:181–192.

696 Ester EF, Anderson DE, Serences JT, Awh E (2013) A Neural Measure of Precision in Visual Working
697 Memory. J Cogn Neurosci 25:754–761.

698 Ester EF, Sprague TC, Serences JT (2015) Parietal and Frontal Cortex Encode Stimulus-Specific
699 Mnemonic Representations during Visual Working Memory. Neuron 87:893–905.

700 Ester EF, Sutterer DW, Serences JT, Awh E (2016) Feature-Selective Attentional Modulations in
701 Human Frontoparietal Cortex. J Neurosci 36:8188–8199.

702 Finn IM, Priebe NJ, Ferster D (2007) The emergence of contrast-invariant orientation tuning in simple
703 cells of cat visual cortex. Neuron 54:137–152.

704 Freeman J, Brouwer GJ, Heeger DJ, Merriam EP (2011) Orientation decoding depends on maps, not
705 columns. J Neurosci 31:4792–4804.

706 Freeman J, Heeger DJ, Merriam EP (2013) Coarse-Scale Biases for Spirals and Orientation in Human
707 Visual Cortex. J Neurosci 33.

708 Furmanski C, Engel S (2000) An oblique effect in human primary visual cortex. Nat Neurosci 3:535–

536.

Fusi S, Miller EK, Rigotti M (2016) Why neurons mix: High dimensionality for higher cognition. Curr Opin Neurobiol 37:66–74.

Gandhi SP, Heeger DJ, Boynton GM (1999) Spatial attention affects brain activity in human primary visual cortex. Proc Natl Acad Sci USA 96:3314–3319.

Garcia JO, Srinivasan R, Serences JT (2013) Near-real-time feature-selective modulations in human cortex. Curr Biol 23:515–522.

Gardner JL (2010) Is cortical vasculature functionally organized? Neuroimage 49:1953–1956.

Gardner JL, Sun P, Waggoner RA, Ueno K, Tanaka K, Cheng K (2005) Contrast adaptation and representation in human early visual cortex. Neuron 47:607–620.

Graf ABA, Kohn A, Jazayeri M, Movshon JA (2011) Decoding the activity of neuronal populations in macaque primary visual cortex. Nat Neurosci 14:239–245.

Hara Y, Gardner JL (2014) Encoding of graded changes in spatial specificity of prior cues in human visual cortex. J Neurophysiol 112:2834–2849.

Hara Y, Pestilli F, Gardner JL (2014) Differing effects of attention in single-units and populations are well predicted by heterogeneous tuning and the normalization model of attention. Front Comput Neurosci 8:12.

Haynes J-D, Rees G (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci 8:686–691.

Ho T, Brown S, van Maanen L, Forstmann BU, Wagenmakers E, Serences JT, Maanen L Van, Forstmann BU, Wagenmakers E, Serences JT (2012) The optimality of sensory processing during

730 the speed-accuracy tradeoff. J Neurosci 32:7992–8003.

731 Hu X, Le TH, Parrish T, Erhard P (1995) Retrospective estimation and correction of physiological
732 fluctuation in functional MRI. Magn Reson Med 34:201–212.

733 Hubel D, Wiesel T (1968) Receptive fields and functional architecture of monkey striate cortex. J
734 Physiol:215–243.

735 Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the
736 cat's visual cortex. J Physiol 160:106–154.2.

737 Itthipuripat S, Garcia JO, Rungratsameetaweemana N, Sprague TC, Serences JT (2014) Changing the
738 spatial scope of attention alters patterns of neural gain in human cortex. J Neurosci 34:112–123.

739 Jehee JFM, Brady DK, Tong F (2011) Attention improves encoding of task relevant features in the
740 human visual cortex. J Neurosci 31:8210–8219.

741 Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nat
742 Neurosci 8:679–685.

743 Kamitani Y, Tong F (2006) Decoding Seen and Attended Motion Directions from Activity in the
744 Human Visual Cortex. Curr Biol 16:1096–1102.

745 Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999) Increased activity in human
746 visual cortex during directed attention in the absence of visual stimulation. Neuron 22:751–761.

747 Kastner S, Ungerleider LG (2000) Mechanisms of visual attention in the human cortex. Annu Rev
748 Neurosci 23:315–341.

749 Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain
750 activity. Nature 452:352–355.

751 Kriegeskorte N, Cusack R, Bandettini P (2010) How does an fMRI voxel sample the neuronal activity
752 pattern: Compact-kernel or complex spatiotemporal filter? Neuroimage 49:1965–1976.

753 Larsson J, Harrison C, Jackson J, Oh S-M, Zeringyte V (2016) Spatial scale and distribution of
754 neurovascular signals underlying decoding of orientation and eye-of-origin from fMRI data. J
755 Neurophysiol 117:jn.00590.2016.

756 Lee J, Maunsell JHR (2010) The effect of attention on neuronal responses to high and low contrast
757 stimuli. J Neurophysiol 104:960–971.

758 Li X, Lu Z-L, Tjan BS, Dosher BA, Chu W (2008) Blood oxygenation level-dependent contrast
759 response functions identify mechanisms of covert attention in early visual areas. Proc Natl Acad
760 Sci U S A 105:6202–6207.

761 Ling S, Jehee JFM, Pestilli F (2015) A review of the mechanisms by which attentional feedback shapes
762 visual selectivity. Brain Struct Funct 220:1237–1250.

763 Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation
764 of the basis of the fMRI signal. Nature 412:150–157.

765 Mante V, Sussillo D, Shenoy K V, Newsome WT (2013) Context-dependent computation by recurrent
766 dynamics in prefrontal cortex. Nature 503:78–84.

767 Martínez-Trujillo JC, Treue S (2002) Attentional modulation strength in cortical area MT depends on
768 stimulus contrast. Neuron 35:365–370.

769 McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single
770 neurons in macaque cortical area V4. J Neurosci Off J Soc Neurosci 19:431–441.

771 Murray SO (2008) The effects of spatial attention in early human visual cortex are stimulus

772 independent. J Vis 8:2.1-11.

773 Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fMRI. Neuroimage
774 56:400–410.

775 Nestares O, Heeger DJ (2000) Robust multiresolution alignment of MRI brain volumes. Magn Reson
776 Med 43:705–715.

777 Norman KA, Polyn SM, Detre GJ, Haxby J V. (2006) Beyond mind-reading: multi-voxel pattern
778 analysis of fMRI data. Trends Cogn Sci 10:424–430.

779 Ogawa S, Lee TM, Kay AR, Tank DW (1990) Brain magnetic resonance imaging with contrast
780 dependent on blood oxygenation. Proc Natl Acad Sci U S A 87:9868–9872.

781 Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K (1992) Intrinsic signal
782 changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance
783 imaging. Proc Natl Acad Sci U S A 89:5951–5955.

784 Olman CA, Ugurbil K, Schrater P, Kersten D (2004) BOLD fMRI and psychophysical measurements
785 of contrast response to broadband images. Vision Res 44:669–683.

786 Pestilli F, Carrasco M, Heeger DJ, Gardner JL (2011) Attentional enhancement via selection and
787 pooling of early sensory responses in human visual cortex. Neuron 72:832–846.

788 Poldrack RA (2006) Can cognitive processes be inferred from neuroimaging data? Trends Cogn Sci
789 10:59–63.

790 Pooresmaeili A, Poort J, Thiele A, Roelfsema PR (2010) Separable codes for attention and luminance
791 contrast in the primary visual cortex. J Neurosci 30:12701–12711.

792 Pratte MS, Sy JL, Swisher JD, Tong F (2016) Radial bias is not necessary for orientation decoding.

793 Neuroimage 127:23–33.

794 Purushothaman G, Bradley DC (2005) Neural population code for fine perceptual decisions in area
795 MT. Nat Neurosci 8:99–106.

796 Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. Annu Rev Neurosci
797 27:611–647.

798 Reynolds JH, Heeger DJ (2009) The Normalization Model of Attention. Neuron 61:168–185.

799 Reynolds JH, Pasternak T, Desimone R (2000) Attention increases sensitivity of V4 neurons. Neuron
800 26:703–714.

801 Sani I, Santandrea E, Morrone MC, Chelazzi L (2017) Temporally evolving gain mechanisms of
802 attention in macaque area V4. J Neurophysiol 118.

803 Saproo S, Serences JT (2010) Spatial attention improves the quality of population codes in human
804 visual cortex. J Neurophysiol 104:885–895.

805 Saproo S, Serences JT (2014) Attention Improves Transfer of Motion Information between V1 and
806 MT. J Neurosci 34:3586–3596.

807 Sasaki Y, Rajimehr R, Kim BW, Ekstrom LB, Vanduffel W, Tootell RBH (2006) The Radial Bias: A
808 Different Slant on Visual Orientation Sensitivity in Human and Nonhuman Primates. Neuron
809 51:661–670.

810 Sclar G, Freeman RD (1982) Orientation selectivity in the cat's striate cortex is invariant with stimulus
811 contrast. Exp Brain Res 46:457–461.

812 Sclar G, Maunsell JH, Lennie P (1990) Coding of image contrast in central visual pathways of the
813 macaque monkey. Vis Res 30:1–10.

814 Scolari M, Byers A, Serences JT (2012) Optimal Deployment of Attentional Gain during Fine
815 Discriminations. J Neurosci 32:7723–7733.

816 Serences JT, Saproo S (2012) Computational advances towards linking BOLD and behavior.
817 Neuropsychologia 50:435–446.

818 Sereno MI, Dale AM, Reppas JB, Kwong KK (1995) Borders of multiple visual areas in humans
819 revealed by functional magnetic resonance imaging. Science (80-) 268:889–893.

820 Shmuel A, Chaimow D, Raddatz G, Ugurbil K, Yacoub E (2010) Mechanisms underlying decoding at
821 7 T: Ocular dominance columns, broad structures, and macroscopic blood vessels in V1 convey
822 information on the stimulated eye. Neuroimage 49:1957–1964.

823 Skottun BC, Bradley A, Sclar G, Ohzawa I, Freeman RD (1987) The effects of contrast on visual
824 orientation and spatial frequency discrimination: a comparison of single cells and behavior. J
825 Neurophysiol 57:773–786.

826 Sprague TC, Saproo S, Serences JT (2015) Visual attention mitigates information loss in small- and
827 large-scale neural codes. Trends Cogn Sci 19:215–226.

828 Sun P, Gardner JL, Costagli M, Ueno K, Allen Waggoner R, Tanaka K, Cheng K (2013)
829 Demonstration of tuning to stimulus orientation in the human visual cortex: A high-resolution
830 fMRI study with a novel continuous and periodic stimulation paradigm. Cereb Cortex 23:1618–
831 1629.

832 Sun P, Ueno K, Waggoner RA, Gardner JL, Tanaka K, Cheng K (2007) A temporal frequency-
833 dependent functional architecture in human V1 revealed by high-resolution fMRI. Nat Neurosci
834 10:1404–1406.

835 Swisher JD, Gatenby JC, Gore JC, Wolfe BA, Moon C-H, Kim S-G, Tong F (2010) Multiscale pattern

836 analysis of orientation-selective activity in the primary visual cortex. J Neurosci 30:325–330.

837 Tootell RB, Hadjikhani NK, Vanduffel W, Liu AK, Mendola JD, Sereno MI, Dale a M (1998)
838 Functional analysis of primary visual cortex (V1) in humans. Proc Natl Acad Sci U S A 95:811–
839 817.

840 Treue S, Maunsell JH (1996) Attentional modulation of visual motion processing in cortical areas MT
841 and MST. Nature 382:539–541.

842 van Bergen RS, Ji Ma W, Pratte MS, Jehee JF (2015) Sensory uncertainty decoded from visual cortex
843 predicts behavior. Nat Neurosci 18:1728–1730.

844 Vintch B, Gardner JL (2014) Cortical correlates of human motion perception biases. J Neurosci
845 34:2592–2604.

846 Wang HX, Merriam EP, Freeman J, Heeger DJ (2014) Motion direction biases and decoding in human
847 visual cortex. J Neurosci 34:12601–12615.

848 Williford T, Maunsell JHR (2006) Effects of spatial attention on contrast response functions in
849 macaque area V4. J Neurophysiol 96:40–54.

850 Yacoub E, Harel N, Ugurbil K (2008) High-field fMRI unveils orientation columns in humans. Proc
851 Natl Acad Sci U S A 105:10607–10612.

852 Yacoub E, Shmuel A, Logothetis N, Uğurbil K (2007) Robust detection of ocular dominance columns
853 in humans using Hahn Spin Echo BOLD functional MRI at 7 Tesla. Neuroimage 37:1161–1177.

854